



Department  
for Transport

# National Travel Survey: Data Extract User Guide, 2002-2012

## Contents

<b>Data Update</b>	<b>2</b>
<b>Key Issues</b>	<b>3</b>
<b>Survey Overview</b>	<b>4</b>
<b>Using NTS Data</b>	<b>7</b>
<b>Key Definitions</b>	<b>8</b>
<b>Weighting</b>	<b>9</b>
<b>Revisions</b>	<b>14</b>
<b>Sample Sizes</b>	<b>15</b>
<b>Annex A – example tabulation</b>	<b>16</b>
<b>Annex B – example tabulation</b>	<b>23</b>
<b>Annex C – example tabulation – Multi-coded data</b>	<b>28</b>

---

## Data Update

### **2002-2012 datasets supersede the 2002-2010 (SN: 5340) files:**

The National Travel Survey (NTS) team in DfT recently completed the redevelopment of the database in which all NTS data is held. All historic data (2002-2007) was copied from a Quantum Quanvert system into a new SQL database. The NTS data from 2008 to 2012 has been fully processed in SQL, and this has over written previous data for years 2008 to 2010 which was previously processed in Quantum and supplied to the UK Data Service.

In addition, as part of the database redevelopment all processing methods were reviewed, including imputation of missing data for some variables. Improvements and corrections have been made which has resulted in some minor revisions to all data published from 2002. See Revision section of this document for details of the main changes made which have affected time series results.

These datasets contain the last set of NTS data covering residents of Great Britain. All data from 2013 onwards will cover residents of England only as the survey coverage changed.

### **All variable names have changed:**

The Table Structures spreadsheet includes a mapping of old to new variable names. This change also affects the variables used for merging files together, see section on Key identifier variables below. The data file containing all trip records is now called Trip.sav; in the 2002-2010 version it was named Journey.sav.

A significant number of extra variables have been supplied to the UK Data Service for the End User Licence dataset. Further additional variables are available to users via Special Licence and the Secure Access Portal. The Table Structures spreadsheet indicates which variables are available and in which survey years – in this document ‘Full’ represents DfT’s own version of the database.

---

## Key issues to consider when analysing NTS data

- The NTS is primarily designed to measure long-term trends in travel in Great Britain and is **not suitable for monitoring short-term trends**.
- **Careful attention should be paid to sample sizes** for all analyses. (See section on Sampling errors)
- **Appropriate 'short walk' weights should be applied to all trip/stage-level analysis** to account for the fact that short walks are only recorded on the seventh day of the travel week. (See section on 'Short walk weighting')
- In addition, there is a weighting strategy to adjust for non-response and 'drop-off', **appropriate weights should be applied to all analyses of data**. (See section on 'Non-response and "drop-off" weighting')
- **All analyses of data from the travel diary should be based on fully co-operating households only**. Analysis at the individual, household and vehicle level should be based on the interview sample which includes data from additional 'partially co-operating' households. (See section on 'Weighting')
- Before carrying out any analysis, all users should **first try to reproduce the examples in the annexes of this document and some published results tables** to ensure they are using the data and applying the weights correctly. The 2012 NTS publication is available at: <https://www.gov.uk/government/publications/national-travel-survey-2012>
- In addition, **users should be familiar with the survey methodology** to ensure data is analysed and interpreted correctly. See details below and the NTS technical information, available at: <https://www.gov.uk/government/collections/national-travel-survey-statistics>
- When analysing the data users should be aware that, although the core survey has remained consistent over time, **some relatively minor changes are made to the survey each year**, e.g. addition/removal of questions, changes to filters and variable categories. The appendix of the annual Technical Reports (see link above) give details of questionnaire changes since 2002.

For advice on using and interpreting the NTS data, please contact:

[national.travelsurvey@dft.gsi.gov.uk](mailto:national.travelsurvey@dft.gsi.gov.uk)

---

## Survey Overview

### Survey background

The NTS is carried out in order to monitor long-term changes in personal travel and provide a better understanding of the use of transport facilities made by different sectors of the population. Individuals in sampled households are interviewed face-to-face to collect personal information, such as age, gender, working status, car access and driving licence holding. They are also asked to complete a seven day travel diary and provide details of trips undertaken, including purpose, method of travel, time of day and trip length.

The first NTS survey was commissioned by the Ministry of Transport in 1965 with further periodic surveys being carried out in 1972/73, 1975/76, 1978/79 and 1985/86. There are a number of differences between the current survey and surveys up to 1978/79. These are detailed in Chapter 7 of the National Travel Survey 2008 Technical Report, which is available on the archived NTS website at:

<http://webarchive.nationalarchives.gov.uk/20091203140650/http://www.dft.gov.uk/pgr/statistics/datatablespublications/personal/methodology/ntstechreports/ntstechrep2008>

### Conducting the survey

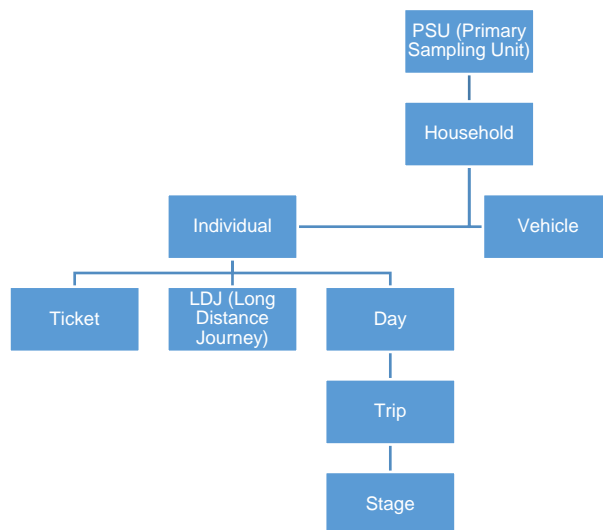
Prior to the interviewer's first call, letters are sent out to the sampled addresses. These introduce the survey and explain that an interviewer will call. The interviewer then arranges an appointment for a 'placement interview'. During this visit, the interviewer obtains information on the household, each household member and on all vehicles to which the household has access, via a computer assisted personal interview (CAPI). The procedures for the seven day travel diary record are then explained. Each household is given a randomly assigned start date for the seven day travel diary.

The placement call is generally followed by a reminder call, just before the start of the travel week, to remind the household to begin their travel records, and by a midweek call during the travel week to check that records are being completed correctly. Within six days of the end of the travel week the interviewer will make a 'pick-up call' to collect the travel records and to check the information recorded with the informants.

### Data types

The data is hierarchical in nature and consists of several record types or 'levels'. Most significantly, data is held about households, vehicles, individuals, trips and stages. Data from each level are linked to make cross-level analysis possible. NTS methodology determines that vehicles belong to households and not individuals. Vehicle data, therefore, can only be analysed by household or vehicle attributes.

## Levels in the NTS database



**PSU** – Primary Sampling Units (PSU) are used to provide a list of areas from which to select a sample of addresses from. In this case each PSU is a postcode sector.

### Key identifier variables

To enable the linking of data from each level of the dataset hierarchy, a set of identifier variables are attached to each data file.

The combination of identifier variables that uniquely identifies a record at each level of the database are:

PSU	- PSUID
Household	- HouseholdID
Vehicle	- VehicleID
Individual	- IndividualID
LDJ	- LDJID
Day	- DayID
Trip	- TripID
Stage	- StageID
Ticket	- IndTicketID

To link levels together the identifier variables from the highest analysis level should match those at the lower level. These identifier variables are new unique IDs which allow for linkage between data levels (without requiring linkage on multiple variables, as in previous versions of the datasets held by the UK Data Service).

---

**Example:**

To link an individual to the trips they made, merge the individual and trip level files and ensure that the individual level identifier, IndividualID, matches on both files:

```
Individual.IndividualID=Trip.IndividualID
```

If merging datasets together, users should remove unwanted variables to reduce the file sizes and length of time it takes to merge.

See examples in Annex A and B for detailed instructions of how to merge data files.

---

## Using NTS Data

### Distance

The length of any trip stage is the distance actually covered, as reported by the traveller, and not the distance as the crow flies. Distance is measured in miles (previous versions of the datasets held by the UK Data Service used tenth of mile).

### Distance by mode

Average distance travelled results are calculated based on the *stage distance* for each mode and are presented as 'average distance travelled by 'mode' – see table [NTS0305](#). Trip rates are usually presented as 'main mode', i.e. based on the mode that is the longest stage distance – see table [NTS0303](#).

### Time

Unit = minutes.

### Denominator

It is standard NTS methodology to include all respondents in the denominator not just those making trips. For example children should be included when calculating driver trips rates.

### Number of records

There are less records in 2011 due to the GPS experiment which was conducted using a sub sample and data collected was not incorporated into the main database.

### Mode: Taxi/minicab

When using main mode or mode variables the split between taxi and minicab should not be used as it is not accurate. Results should be presented combined like published tables, e.g. [NTS0303](#)

### Mode: Walking

Short walks (those under one mile) are only collected on day 7 of the travel diary. 'Short walk weights' account for this and should be applied when producing analysis – see 'Weighting' section below. For detailed walking trip analysis it may be necessary to limit the base sample to day 7 only data, e.g. [NTS9909](#).

---

## Key Definitions

### Trip

A **trip** is defined as a one-way course of travel having a single main purpose, e.g. a walk to school or a trip to work without any break in travel.

Usually the respondent is clear what the single main purpose of a particular trip is. However, sometimes people go out for a number of reasons, or go out for one main reason but carry out a number of different activities, perhaps at different places. Complex travel like this is broken into separate trips so that the data can be analysed. Where a stop is entirely secondary to the main purpose (such as a stop to buy a newspaper on the way to work), the stop is disregarded.

### Stage

A trip consists of one or more **stages**. A new stage is defined when there is a change in the form of transport or when there is a change of vehicle requiring a separate ticket.

In order to reduce the burden on respondents, travel involving a number of stops for the same main purpose and using the same form of transport are treated as one continuous 'series of calls' trip from the first such call to the last one. Only shopping and 'in course of work' travel can be treated in this way. A doctor's round could therefore consist of one trip to the first patient, one series of calls trip to the other patients and one trip from the last call back to the surgery or home.

A full list of definitions is available in the 2012 Notes and Definitions document available at: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/226817/nts2012-notes.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/226817/nts2012-notes.pdf)



---

## Weighting

### Short walk weighting

Because trips of less than one mile in distance are recorded only on the seventh day of the travel week, these trips must be weighted by a factor of seven when analysed. Also, for consistency with earlier surveys, 'series of calls' trips are excluded from analysis of stage and trip counts and time. Therefore, one of the following 'short walk weights' must be applied to any tabulations using trip or stage counts, distance or time:

<b>SSXSC</b>	Number of stages to be counted, grossed for short walks and excluding 'series of calls' trips.
<b>SD</b>	Stage distance travelled, grossed for short walks.
<b>STTXSC</b>	Stage travelling time grossed for short walks and excluding 'series of calls' trips.
<b>JJXSC</b>	Number of trips to be counted, grossed for short walks and excluding 'series of calls' trips.
<b>JD</b>	Trip distance travelled, grossed for short walks.
<b>JOTXSC</b>	Overall trip time (includes travelling and waiting time), grossed for short walks and excluding 'series of calls' trips.
<b>JTTXSC</b>	Trip travelling time, grossed for short walks and excluding 'series of calls'.

These weighted variables have been constructed as follows:

### SSXSC

If 'series of calls'	SSXSC = 0
If not 'series of calls' and 'short walk stage'	SSXSC = 7
If not 'series of calls' and not 'short walk stage'	SSXSC = 1

---

## STTXSC

If 'series of calls'	$STTXSC = 0 * StageTime$
If not 'series of calls' and 'short walk stage'	$STTXSC = 7 * StageTime$
If not 'series of calls' and not 'short walk stage'	$STTXSC = 1 * StageTime$

## SD

If 'short walk stage'	$SD = 7 * StageDistance$
If not 'short walk stage'	$SD = 1 * StageDistance$

## JJXSC

If 'series of calls'	$JJXSC = 0$
If not 'series of calls' and 'short walk trip'	$JJXSC = 7$
If not 'series of calls' and not 'short walk trip'	$JJXSC = 1$

## JOTXSC

If 'series of calls'	$JOTXSC = 0 * TripTotalTime$
If not 'series of calls' and 'short walk trip'	$JOTXSC = 7 * TripTotalTime$
If not 'series of calls' and not 'short walk trip'	$JOTXSC = 1 * TripTotalTime$

## JTTXSC

If 'series of calls'	$JTTXSC = 0 * TripTravTime$
If not 'series of calls' and 'short walk trip'	$JTTXSC = 7 * TripTravTime$
If not 'series of calls' and not 'short walk trip'	$JTTXSC = 1 * TripTravTime$

## JD

$JD = \text{Sum}(SD) \text{ Stage distances}$
---

---

## Non-response and "drop-off" weighting

In addition to the short walk weights described above, a weighting strategy for the NTS was developed following a recommendation in the 2000 National Statistics Quality Review of the NTS. For the first time, the 2005 NTS results were based on weighted data. The weighting methodology was applied to data back to 1995 and all NTS figures for 1995 onwards which are published or released are now based on weighted data.

As well as adjusting for non-response bias, the weighting strategy for the NTS also adjusts for the drop-off in the number of trips recorded by respondents during the course of the travel week; for uneven recording of short walks by day of the week and for the short-fall in reporting of long distance trips.

Further information on the weighting methodology is available in the 2012 Technical Report at the link below. Please note that "W" notation in the Technical Report is not the same as the "W" notation in the NTS datasets as described in the 'Weighting matrices' section below).

[https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/225735/nts2012-technical.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/225735/nts2012-technical.pdf)

**Therefore, there are several sets of weights which apply to different levels of the database;** household, trip and long distance journey. It is important to select the correct weights for each analysis. Initial results should be checked against published data to ensure weights are being applied correctly.

There are two samples which can be used for analysis:

### Diary sample

Analysis of travel data is based on the diary sample. This comprises all 'fully co-operating households', defined as households for which the following information is available: a household interview, an individual interview for each household member, a seven day travel diary for each individual and, where applicable, at least one completed vehicle section. Weights were produced to adjust for non-response and, at the trip-level, they were also produced for drop-off in recording observed during the seven day travel week.

### Interview sample

Analyses at household, individual and vehicle level are based on the interview sample. This sample comprises all fully co-operating households included in the diary sample, together with some additional 'partially co-operating households'. Generally these partially co-operating households had co-operated fully with the various interviews but not all household members had completed the travel diary. Prior to the introduction of the weighting methodology, data from partially co-operating households was not included in NTS analyses but this expanded dataset is now used for analyses which do not require data from the seven day travel record.

Fully co-operating households (Interview data + travel data)
Partially co-operating households (Interview data only)

} Diary sample }

It is important to use the correct sample for all analysis.

The weighting methodology produces weights at the household, LDJ and trip level. The household weights apply to all individuals and vehicles within the household, and they have therefore been attached to the individual and vehicle files for ease of use. Similarly, the trip level weights apply to all stages within trips and have therefore been attached to the stage-level records for ease of use.

The variable ' OutCom\_B02ID' (1=fully productive, 0=partially productive) is included on the household file to enable users to select the diary sample. This variable can be merged into other files if necessary, e.g. individual file.

## Weighting matrices

The following weighting matrices are available and should be used on the appropriate level data for the analysis being produced:

### W1

**Unweighted diary sample** - this gives unweighted results for the diary sample only. (This is equivalent to the results produced before the weighting strategy was introduced and can be used to generate unweighted sample sizes for analysis of the diary sample. It is effectively the same as the OutCom\_B02ID variable mentioned above – formerly named the ‘status’ variable).

### W2

**Diary sample household weight** – (adjusts for non-response) apply to all analysis of the diary sample at **household**, **individual** and **vehicle** level.

### W3

**Interview sample household weight** - apply to all analysis of the interview sample at **household**, **individual** and **vehicle** level.

### W4

**LDJ weight incorporating household weight** - apply to all analysis at **LDJ** level

---

## W4xhh

**LDJ weight excluding household weight** – apply only if also applying a household weight (i.e. W2 or W3)

## W5

**Trip/stage weight** - apply to all analysis of **trip** and **stage** data

## W5xhh

**Trip/stage weight excluding household weight** - apply only if also applying a household weight (i.e. W2 or W3)

## No weighting matrix or W0

If no weighting matrix is applied, this gives unweighted results for the interview sample.

- For most analyses at household, individual and vehicle level, w3 should be applied.
- For most analyses of travel patterns, w5 should be applied to trip/stage data and w2 should be applied at the individual level in order to calculate rates. In addition, the short walk weights should be applied (see 'Short walk weighting' above).

## Examples of applying weights:

- To generate trip rates - apply w5 to trip data and apply w2 to individual data (i.e. Diary sample)
- To calculate household car ownership - apply w3 to the household data (Interview sample)
- To calculate the proportion of driving licence holders - apply w3 to the individual data (Interview sample)
- To determine the unweighted sample size for trip rate analysis - apply w1 to the trip data and w1 to the individual data (Diary sample)
- To determine the unweighted sample size for household car ownership or driving licence figures - apply no weights or w0 (Interview sample)

## Long distance journey (LDJ) weights

Long distance journeys (trips over 50 miles) are collected in two ways in the NTS – via the travel diary and in the interview.

From 2002-2005 long distance trips were collected during the interview for an additional period of 3 weeks prior to the start of the travel diary. From 2006 onwards, they have been collected during the single week prior to the placement interview.

Analysis of the LDJ data a few years ago highlighted concerns over the quality of the interview recall data compared to the data collected during the travel week via the diary. Therefore the LDJ weighting methodology was revised back to 2006 to take account of the

---

systematic under-reporting of shorter long distance journeys. The re-weighting had little impact upon the overall LDJ trip rates. However, it resulted in a downward adjustment to distance estimates.

## Revisions

The following are the main changes made to the 2002-2012 datasets which account for the difference in the datasets for 2002-2010 previously available from the UK Data Service.

### SQL processing of 2008-2012 raw data

Due to changed imputation methodologies for key variables, e.g. trip mode and purpose, the trip data processed in SQL for 2008-2010 needed to be re-weighted. In most cases the overall results have changed very little from those processed in the old database.

### Geographic data

A full recode of all geographic variables at all levels was conducted so that there was a consistent time series available from 2002-2012 at household level. This included creating a new Area Type variable which now correctly allocates households in South Yorkshire to the Metropolitan built-up areas category. This recoding also now makes the allocation of Scottish postcodes to an area type prior to 2007 consistent with 2007 onwards. More information on Area type can be found in the 2012 [Notes & Definitions](#) document. All published results tables by region and area type are now all based on household geographies rather than at the PSU level so there are some minor differences.

### Ticket data

The data on tickets used and the associated costs for each stage have been re-processed in a consistent way for all years from 2002. These data are now available in a separate ticket file – previously they were within the individual data file.

### Corrections

Whilst working on the database redevelopment project some errors were found in the processing of some NTS data in Quantum Quanvert. These errors have been corrected in the new SQL database and therefore the 2002-2012 raw datasets. The allocation of 'missing' (NA) and 'does not apply' (DNA) cases has also been made consistent over all years.

### Imputation indicators

The NTS now includes imputation indicators in the data files for Household, Vehicle, Individual, Trip and Stage for some key variables from 2008 to 2012. These indicators inform users if the value was originally missing and hence imputed.

# Weighted and unweighted sample sizes, 1995-2012

Sample numbers on which analyses are based: Great Britain, 1995 to 2012<sup>1</sup>

	1995	1996	1997	1998	1999	2000	2001	2002 <sup>2</sup>	2003 <sup>2</sup>	2004 <sup>2</sup>	2005 <sup>2</sup>	2006 <sup>2</sup>	2007 <sup>2</sup>	2008 <sup>2</sup>	2009 <sup>2</sup>	2010 <sup>2</sup>	2011 <sup>2</sup>	2012
<b>Household sample</b>																		
Unweighted diary	3,211	3,210	3,139	2,935	3,020	3,435	3,469	7,437	8,258	8,122	8,430	8,297	8,431	8,094	8,384	8,097	7,741	8,201
Unweighted interview	3,491	3,505	3,465	3,300	3,376	3,783	3,760	8,849	9,196	8,991	9,453	9,261	9,278	8,924	9,128	8,775	8,461	8,972
Weighted diary	3,209	3,211	3,139	2,938	3,018	3,431	3,472	7,437	8,258	8,122	8,430	8,297	8,431	8,094	8,384	8,097	7,741	8,201
Weighted interview	3,490	3,506	3,463	3,303	3,374	3,782	3,761	8,849	9,196	8,991	9,453	9,261	9,278	8,924	9,128	8,775	8,461	8,972
<b>Individual sample</b>																		
Unweighted diary	7,723	7,665	7,473	6,842	6,970	8,056	7,978	16,886	19,467	19,199	19,904	19,490	19,735	18,983	19,914	19,072	18,069	19,154
Unweighted interview	8,521	8,504	8,452	7,945	8,000	9,054	8,833	20,827	21,990	21,588	22,702	22,141	21,931	21,165	21,835	20,839	19,988	21,243
Weighted diary	7,675	7,704	7,487	6,986	7,109	8,114	8,132	17,494	19,578	19,302	20,103	19,794	19,940	19,253	19,918	19,210	18,339	19,547
Weighted interview	8,348	8,405	8,258	7,857	7,945	8,942	8,811	20,789	21,795	21,369	22,539	22,098	21,959	21,219	21,706	20,819	20,057	21,385
<b>Child (&lt;16 yrs) sample</b>																		
Unweighted diary	1,744	1,666	1,650	1,433	1,466	1,731	1,659	3,413	4,178	4,129	4,150	3,938	3,971	3,804	4,042	3,865	3,540	3,661
Unweighted interview	1,945	1,869	1,917	1,708	1,703	1,971	1,860	4,398	4,702	4,690	4,776	4,497	4,408	4,260	4,414	4,193	3,921	4,115
Weighted diary	1,602	1,634	1,547	1,447	1,467	1,666	1,624	3,523	3,900	3,815	3,963	3,848	3,759	3,684	3,790	3,660	3,524	3,718
Weighted interview	1,741	1,782	1,704	1,626	1,631	1,835	1,764	4,181	4,336	4,222	4,443	4,296	4,141	4,057	4,118	3,962	3,852	4,060
<b>Adult (16+) sample</b>																		
Unweighted diary	5,979	5,999	5,823	5,409	5,504	6,325	6,319	13,473	15,289	15,070	15,754	15,552	15,764	15,179	15,872	15,207	14,529	15,493
Unweighted interview	6,576	6,635	6,535	6,237	6,297	7,083	6,973	16,429	17,288	16,898	17,926	17,644	17,523	16,905	17,421	16,646	16,067	17,128
Weighted diary	6,074	6,069	5,940	5,539	5,640	6,447	6,508	13,971	15,678	15,487	16,141	15,945	16,181	15,569	16,128	15,550	14,815	15,829
Weighted interview	6,606	6,623	6,553	6,231	6,315	7,107	7,047	16,608	17,459	17,147	18,097	17,801	17,818	17,162	17,587	16,857	16,206	17,325
<b>Motor vehicle sample</b>																		
Unweighted diary	3,296	3,301	3,238	3,121	3,217	3,772	3,707	8,195	9,264	9,065	9,847	9,758	9,898	9,511	9,985	9,750	9,192	9,702
Unweighted interview	3,642	3,691	3,653	3,608	3,681	4,240	4,081	9,954	10,452	10,190	11,228	11,118	10,975	10,588	10,911	10,675	10,136	10,678
Weighted diary	3,339	3,374	3,313	3,201	3,270	3,843	3,780	8,391	9,408	9,261	10,059	9,875	10,005	9,655	9,956	9,790	9,214	9,664
Weighted interview	3,629	3,694	3,663	3,601	3,664	4,237	4,090	9,959	10,465	10,270	11,264	11,028	10,995	10,636	10,841	10,594	10,079	10,569
<b>Trip sample</b>																		
Unweighted diary	126,088	124,748	122,397	112,867	114,501	130,179	129,998	278,916	314,845	310,065	322,500	312,347	305,077	295,791	306,743	290,803	272,208	289,573
Unweighted diary (short walks grossed up & excluding series of calls)	156,729	153,366	150,403	137,851	138,233	155,644	155,789	326,495	369,664	363,692	379,315	369,072	353,191	341,529	358,500	337,402	316,494	335,288
Unweighted interview	129,356	133,896	127,242	120,996	123,182	137,689	139,240	302,796	333,833	326,869	345,996	336,802	324,882	319,902	320,257	308,431	291,072	310,732
Weighted diary (short walks grossed up & excluding series of calls)	159,163	162,418	154,703	146,148	147,271	162,950	165,284	351,123	388,152	379,887	402,356	393,812	371,823	366,326	371,473	354,900	337,029	357,616
Weighted interview																		
<b>Stage sample</b>																		
Unweighted diary	131,548	129,690	127,273	117,269	119,072	136,324	134,036	289,048	327,230	322,602	335,940	326,076	318,996	309,722	320,858	304,942	284,718	303,436
Unweighted diary (short walks grossed up & excluding series of calls)	174,261	169,384	166,421	150,833	151,978	172,463	167,939	349,251	396,761	391,871	408,775	399,973	384,402	371,642	390,741	370,279	345,030	367,019
Unweighted interview																		
Weighted diary	135,017	139,423	132,494	125,838	128,346	144,406	143,953	314,728	348,024	341,321	361,449	352,392	340,520	335,877	335,727	324,118	305,012	326,477
Weighted diary (short walks grossed up & excluding series of calls)	177,225	180,055	172,061	160,462	162,999	181,506	179,767	378,777	419,996	412,966	437,211	430,032	407,550	401,690	408,304	392,572	370,194	395,017
Weighted interview																		

<sup>1</sup> Data for 1995 to 2001 is based on calendar years. Data for 2002 to 2012 is based on survey years which run mid-January to mid-January.

[https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/225736/nts2012-sample.xls](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/225736/nts2012-sample.xls)

**Note:** Between 1995 and 2001 published NTS analysis is based on the calendar year in which a household started completing their travel diary. Analysis from 2002 onwards is based on the survey year (mid-January to mid-January) in which the household was sampled.

---

## Annex A: Example tabulation 1

NTS data users should replicate analysis produced by DfT to ensure that they understand the methodology of weighting and the grossing up of short walks.

### Trips per person per year by gender and survey year

Load the journey level SPSS data file 'Trip.sav'

Load the individual level SPSS data file 'Individual.sav'

Before performing the analysis you should ensure that both files are sorted based on the key identifier variables at each level.

Select the **Trip** file.

From the menu select **Data -> Sort Cases**

Highlight **TripID** and move over to the 'Sort by' list. Ensure that **ascending** is checked.

The 'Sort by' list should contain:

TripID (A)

Select OK.

Select the **Individual** file.

From the menu select **Data -> Sort Cases**

Highlighting **IndividualID** and move over to the 'Sort by' list. Ensure that ascending is checked.

The 'Sort by' list should contain

IndividualID (A)

Select OK.



---

To disaggregate the data by the gender of the individual making the trip, individual level data needs to be attached to the trip file.

Select the **Trip** file.

From the menu **select Data -> Merge files -> Add variables**

Select '**An open dataset**' and then select file '**Individual.sav**'

Select continue.

Data should be linked using the identifier variable at individual level – **IndividualID**.

Check '**Match cases on key variables**' and '**Cases are sorted in order of key variables in both datasets**'

Check '**Non-active dataset is keyed table**'

From the excluded variables list highlight IndividualID and move to the key variables box

Select OK.

A message will appear 'Warning: Keyed match will fail if data are not sorted in ascending order of key variables'

This can be ignored as we have already ensured the correct sorting order.

The individual characteristics of the person making each trip will now be attached to the journey data.

Before producing the tabulation of gender by survey year, the weighting must be specified:

From the menu select Data -> Weight cases

Check 'Weight cases by' and move across W5 (trip weight) as the frequency variable.

Select OK.

---

The tabulation can now be specified.

From the menu select **Analyze -> Compare Means -> Means**

To gross up short walks we must use variable **JJXSC** as the dependent list variable.

Move JJXSC across to 'Dependent List'

Select **Gender** as the layer 1 variable by moving **Sex\_B01ID** across to the independent list selection

Select 'Next' to choose the 2<sup>nd</sup> layer variable

Select **Survey Year** as the layer 2 variable by moving **SurveyYear** across to the independent list selection

Select '**Options**' to determine the type of statistics to perform.

Move '**Sum**' across to the 'Cell statistics' box. Remove all other statistics.

Select continue.

Select OK.

The following SPSS output should be produced:

**Report**

Sum

Sex of person	Survey year - actual year	Number of trips - grossing up short walks and excluding series of calls
Male	2002	170087
	2003	188735
	2004	183612
	2005	194585
	2006	188663
	2007	177980
	2008	175774
	2009	177365
	2010	169684
	2011	161512
	2012	172216
	Total	1960213
Female	2002	181036
	2003	199417
	2004	196275
	2005	207770
	2006	205150
	2007	193844
	2008	190552
	2009	194107
	2010	185215
	2011	175516
	2012	185400
	Total	2114284
Total	2002	351123
	2003	388152
	2004	379887
	2005	402356
	2006	393812
	2007	371823
	2008	366326
	2009	371473
	2010	354900
	2011	337029
	2012	357616
	Total	4074497

These are the weighted and grossed up number of trips *per week* for each gender/survey year combination.

For each gender/survey year the weighted number of individuals also needs to be known.  
Note – it is standard NTS methodology to include all respondents in the denominator not just those making a trip.

Select the '**Individual.sav**' file.

Before producing the tabulation of gender by survey year, the weighting must be specified

From the menu select **Data -> Weight cases**

Check '**Weight cases by**' and move across **W2** (household non-response weight) as the frequency variable.

Select OK.

The tabulation can now be specified.

From the menu select '**Analyze**' -> '**Descriptive Statistics**' -> '**Crosstabs**'

Move **Survey Year** across to the row specification box

Move **Sex\_B01ID** across to the column specification box

Select **Cells** and in the **Non-integer weights** section select '**No adjustments**'

Select OK.

The following SPSS output should be produced:

Survey year - actual year \* Sex of person Crosstabulation

Count

		Sex of person		Total
		Male	Female	
Survey year - actual year	2002	8549.647	8944.077	17493.723
	2003	9578.418	9999.596	19578.014
	2004	9451.343	9850.284	19301.626
	2005	9843.876	10259.522	20103.397
	2006	9699.779	10093.774	19793.553
	2007	9772.677	10167.305	19939.982
	2008	9454.678	9798.554	19253.233
	2009	9770.212	10147.339	19917.551
	2010	9435.616	9774.031	19209.647
	2011	9008.588	9330.581	18339.169
	2012	9604.327	9942.373	19546.700
	Total	104169.160	108307.436	212476.596

The trip counts can now be divided by the individual sample. It should be noted that the trip counts are for a single week so therefore need to be multiplied by 52.14 (number of weeks in a year).

$$\begin{aligned} \text{e.g. Trips per male in 2002} &= \frac{170,087}{8,549.647} \times 52.14 \\ &= 1,037 \end{aligned}$$

The following results should be calculated showing trips per person per year by gender:

	Year	Trip rate
<b>Male</b>	2002	1037
	2003	1027
	2004	1013
	2005	1031
	2006	1014
	2007	950
	2008	969
	2009	947
	2010	938
	2011	935
	2012	935
	2002/12 average	981
<b>Female</b>	2002	1055
	2003	1040
	2004	1039
	2005	1056
	2006	1060
	2007	994
	2008	1014
	2009	997
	2010	988
	2011	981
	2012	972
	2002/12 average	1018
<b>Total</b>	2002	1047
	2003	1034
	2004	1026
	2005	1044
	2006	1037
	2007	972
	2008	992
	2009	972
	2010	963
	2011	958
	2012	954
	2002/12 average	1000

---

Finally, repeat the above step using W1 to get the unweighted number of individuals. This needs to be known to ensure that the base sample sizes are sufficient.

Select the '**Individual.sav**' file.

Before producing the tabulation of gender by survey year, the weighting must be specified

From the menu select **Data -> Weight cases**

Check '**Weight cases by**' and move across **W1** (unweighted diary weight) as the frequency variable.

Select OK.

The tabulation can now be specified.

From the menu select '**Analyze**' -> '**Descriptive Statistics**' -> '**Crosstabs**'

Move **Survey Year** across to the row specification box

Move **Sex\_B01ID** across to the column specification box

Select OK.

The following SPSS output should be produced:

### Survey year - actual year \* Sex of person Crosstabulation

Count

		Sex of person		Total
		Male	Female	
Survey year - actual year	2002	8085.000	8801.000	16886.000
	2003	9360.000	10107.000	19467.000
	2004	9219.000	9980.000	19199.000
	2005	9560.000	10344.000	19904.000
	2006	9396.000	10094.000	19490.000
	2007	9490.000	10245.000	19735.000
	2008	9127.000	9856.000	18983.000
	2009	9583.000	10331.000	19914.000
	2010	9232.000	9840.000	19072.000
	2011	8718.000	9351.000	18069.000
	2012	9244.000	9910.000	19154.000
	Total	101014.000	108859.000	209873.000

## Annex B: Example tabulation 2

### Miles per person per year, by mode (at stage level)

Load the stage level data file 'Stage.sav'

Load the individual level data file 'Individual.sav'

Before performing the analysis you should ensure that both files are sorted based on the key identifier variables at each level.

Select the **Stage** file.

From the menu select **Data -> Sort Cases**

Highlight **StageID** and move over to the 'Sort by' list. Ensure that ascending is checked.

Select OK.

Select the **Individual** file.

From the menu select **Data -> Sort Cases**

Highlighting **IndividualID** and move over to the 'Sort by' list. Ensure that ascending is checked.

Select OK.

---

Before producing the tabulation of stage distance by survey year and mode, the weighting must be specified.

Select the **Stage** file

From the menu select **Data -> Weight cases**

Check '**Weight cases by**' and move across **W5** (stage weight) as the frequency variable.

Select OK.

The tabulation can now be specified.

From the menu select **Analyze -> Compare Means -> Means**

To gross up short walks we must use variable **SD** as the dependent list variable.

Move **SD** across to 'Dependent List'

Select **SurveyYear** as the layer 1 variable by moving it across to the independent list selection

Select 'Next' to choose the 2<sup>nd</sup> layer variable

Select **Stage Mode** as the layer 2 variable by moving **StageMode\_B01ID** across to the independent list selection

Select '**Options**' to determine the type of statistics to perform.

Move '**Sum**' across to the '**Cell statistics**' box. Remove all other statistics.

Select continue.

Select OK.



The following output should be produced (for all years down to 2012):

Report		
Sum		
Survey year - actual year	Stage mode of travel - as recorded in diary - 18 categories	Stage distance - grossing up short walks
2002	Walk	66278.50
	Bicycle	12132.08
	Private (hire) bus	41637.52
	Car	1845539.78
	Motorcycle, scooter, moped	11629.80
	Van/lorry	92132.81
	Other private	6914.02
	London stage bus	18742.35
	Other stage bus	75319.03
	Coach/Express bus	10923.44
	Excursion/Tour bus	9005.67
	London Underground	27065.54
	Surface Rail	138536.02
	Light rail	2536.75
	Air	14752.25
	Taxi	13266.63
	Minicab	6594.07
	Other public	1073.24
	Total	2394079.50

These are the weighted and grossed up number of stage distances *per week* for each survey year/mode combination.

For each mode/survey year the number of individuals also needs to be known. Note – it is standard NTS methodology to include all respondents in the denominator not just those making a trip.

Select the '**Individual.sav**' file.

From the menu select **Data -> Weight cases**

Check '**Weight cases by**' and move across **W2** (household weight) as the frequency variable.

Select OK.

Before producing the tabulation of mode by survey year, the weighting must be specified.

From the menu select '**Analyze**' -> '**Descriptive Statistics**' -> '**Frequencies**'

Move **Survey Year** across to the Column box

Select OK.

The tabulation can now be specified.

The following output should be produced:

**Survey year - actual year**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2002	17494	8.2	8.2	8.2
	2003	19578	9.2	9.2	17.4
	2004	19302	9.1	9.1	26.5
	2005	20103	9.5	9.5	36.0
	2006	19794	9.3	9.3	45.3
	2007	19940	9.4	9.4	54.7
	2008	19253	9.1	9.1	63.8
	2009	19918	9.4	9.4	73.1
	2010	19210	9.0	9.0	82.2
	2011	18339	8.6	8.6	90.8
	2012	19547	9.2	9.2	100.0
	Total	212477	100.0	100.0	

---

The stage distances can now be divided by the individual sample (the frequency column). It should be noted that the stage distances are for a single week so therefore need to be multiplied by 52.14 (number of weeks in a year).

$$\begin{aligned}\text{e.g. Trips per male in 2002} &= \frac{66278.50}{17,494} \times 52.14 \\ &= 197.54\end{aligned}$$

As in Annex A, repeat the above step using W1 to get the unweighted number of individuals. This needs to be known to ensure that the base sample sizes are sufficient.

Select the '**Individual.sav**' file.

From the menu select **Data -> Weight cases**

Check '**Weight cases by**' and move across **W1** (unweighted diary weight) as the frequency variable.

Select OK.

The tabulation can now be specified.

From the menu select '**Analyze**' -> '**Descriptive Statistics**' -> '**Frequencies**'

Move **Survey Year** across to the Column box

Select OK.

The following SPSS output should be produced:

---

Survey year - actual year

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2002	16886	8.0	8.0	8.0
	2003	19467	9.3	9.3	17.3
	2004	19199	9.1	9.1	26.5
	2005	19904	9.5	9.5	36.0
	2006	19490	9.3	9.3	45.2
	2007	19735	9.4	9.4	54.6
	2008	18983	9.0	9.0	63.7
	2009	19914	9.5	9.5	73.2
	2010	19072	9.1	9.1	82.3
	2011	18069	8.6	8.6	90.9
	2012	19154	9.1	9.1	100.0
	Total	209873	100.0	100.0	

In addition, the unweighted number of stages for each mode should be obtained to ensure that the sample sizes on which the miles are based are sufficient.

## Annex C: Example tabulation 3 – Multi-coded data

### Reasons for not learning to drive by age

Load the individual level data file 'Individual.sav'

Before producing the tabulation of individuals and their reasons for not learning to drive by age, the weighting must be specified.

---

From the menu select **Data -> Weight cases**

Check '**Weight cases by**' and move across **W3** (interview sample weight) as the frequency variable.

Select OK.

In this case we are only interested in data for 2012. To restrict the data file to cases for 2012

From the menu select **Data -> Select cases**

Check '**If condition is satisfied**' and select **If...**

Move **SurveyYear** across to the definition box and enter '**=2012**'.

The box should read '**SurveyYear=2012**'.

Select OK.

only:

Any outputs will now only use individuals from the 2012 survey.

The tabulation can now be specified.

From the menu select '**Analyze**' -> '**Descriptive Statistics**' -> '**Crosstabs**'

Move **ResNDN\_A\_B01ID** across to the row specification box

Repeat this step for all **ResNDN** variables (B to P).

Move **Age\_B04ID** across to the column specification box

Select **Cells** and in the **Non-integer weights** section select '**No adjustments**'

Select OK.

The following SPSS output should be produced (for each response variable – there are 16 possible responses, A-P):

Reason individual does not drive - Family or friends can drive me when necessary \* Age of person - banded age - Band D - All ages - 9 categories \* Survey year - actual year Crosstabulation

Count

Survey year - actual year			Age of person - banded age - Band D - All ages - 9 categories									Total
			0 - 4 years	5 - 10 years	11 - 16 years	17 - 20 years	21 - 29 years	30 - 39 years	40 - 49 years	50 - 59 years	60 years +	
2012	Reason individual does not drive - Family or friends can drive me when necessary	Yes	.000	.000	.000	132.966	166.050	133.267	111.638	134.147	417.959	1096.027
		No	1348.569	1454.680	1525.500	910.225	2419.969	2700.392	3063.582	2520.721	4345.085	20288.725
	Total		1348.569	1454.680	1525.500	1043.191	2586.019	2833.659	3175.221	2654.868	4763.045	21384.752
Total	Reason individual does not drive - Family or friends can drive me when necessary	Yes	.000	.000	.000	132.966	166.050	133.267	111.638	134.147	417.959	1096.027
		No	1348.569	1454.680	1525.500	910.225	2419.969	2700.392	3063.582	2520.721	4345.085	20288.725
	Total		1348.569	1454.680	1525.500	1043.191	2586.019	2833.659	3175.221	2654.868	4763.045	21384.752

The weighted number of individuals also needs to be known.

From the menu select **Data -> Weight cases**

Check '**Do not weight cases**'

Select OK.

First, turn the weighting off:

Next, we must ensure that we are only counting individuals who have been asked their

From the menu select **Data -> Select cases**

Select **If...** condition.

Edit the definition to select individuals answering the question.

The box should read '**SurveyYear=2012 & (ResNDN\_NA\_B01ID=2 & ResNDN\_DNA\_B01ID=2 & ResNDN\_DEAD\_B01ID=2)**'.

Select OK.

reasons for not learning to drive.

From the menu select **Analyze -> Compare Means -> Means**

To count the weighted number of individuals we need to count **W3**.

Move **W3** across to 'Dependent List'

Select **Age\_B04ID** as the layer 1 variable by moving it across to the independent list selection

Select '**Options**' to determine the type of statistics to perform.

Move '**Sum**' across to the '**Cell statistics**' box. Remove all other statistics.

Select continue.

Select OK.

The tabulation can now be specified:

The following SPSS output should be produced:

**Report**

Sum

Age of person - banded age - Band D - All ages - 9 categories	Weighted interview sample
17 - 20 years	447.1548920
21 - 29 years	696.8862122
30 - 39 years	531.8084065
40 - 49 years	449.5771186
50 - 59 years	470.9010953
60 years +	1161.584938
Total	3757.912662

To calculate all percentages (as in [NTS0203](#)) divide the weighted number of individuals for each response and age group by the relevant total number of individuals.

e.g. Percentage of individuals aged 17-20 with the reason 'Family or friends can drive me when necessary' given =

$$\frac{132.9655693}{447.1548920} \times 100 = 30\%$$

Repeat, the above step with **W0** (no weight) to get the unweighted number of individuals.

---

The following SPSS output should be produced:

**Report**

Sum

Age of person - banded age - Band D - All ages - 9 categories	Unweighted interview sample
17 - 20 years	429
21 - 29 years	585
30 - 39 years	465
40 - 49 years	401
50 - 59 years	470
60 years +	1240
Total	3590